MULTIWAY DATA ANALYSIS
R. Coppi and S. Bolasco (Editors)

SINGULAR VALUE DECOMPOSITIONS OF INTERACTIONS IN THREE-WAY CONTINGENCY TABLES

Pieter M. KROONENBERG

Department of Education, University of Leiden & Department of Psychology, University of Queensland
Department of Education, University of Leiden, P. O. Box 9700, 2300 RA Leiden, The Netherlands

In this paper generalizations of the singular value decomposition are used to analyze interactions from three-way contingency tables. These decompositions are primarily applied to standardized residuals from various loglinear models to produce three-way generalizations of correspondence analysis

1. INTRODUCTION

Contingency tables turn up in many research projects in many contexts, and there exists an extensive collection of techniques for their analysis. Especially in recent years the development of loglinear models for contingency table analysis has enabled researchers to make more detailed statements about association in multi-way tables than just reporting descriptive levels of significance. Notwithstanding the refined machinery connected with loglinear models, there are problems with their application to large tables and/or higher-dimensional tables. Two of these problems are the difficulty of interpreting interaction terms when there are many of them (as is the case with large tables), and the complexity of interpreting higher-order interaction terms, especially if there are many observations. In this paper we will look at these problems for three-way contingency tables. Although extensions to higher-way tables are possible, we will not consider these here.

2. LOGLINEAR MODELS AND STANDARDIZED RESIDUALS

A saturated model for the expected values in a contingency table is a model which completely accounts for the data by specifying all effects and interactions. For a two-way table it has the form, using Bishop, Fienberg, & Holland's (1975) notation (p. 17ff.)

$$\log e_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \qquad i=1,..,I;\ j=1,..,J,$$

with e_{ij} the expected cell count. There are two main effects vectors $\mathbf{u}_1$ and $\mathbf{u}_2$, and one two-way interactions matrix $\mathbf{u}_{12}$. A non-saturated model consists of a combination of at most two of the terms at the right hand side. The most common model for a two-way table is the model of independence between rows and columns

$$\log e_{ij} = u + u_{1(i)} + u_{2(j)},$$

This model may be tested against the null hypothesis of independence via Pearson's X^2-test

$$X^2 = \sum (f_{ij} - e_{ij})^2/e_{ij},$$

or the likelihood ratio statistic

$$G^2 = 2 \sum f_{ij} \log(f_{ij}/e_{ij}).$$

In this paper we will exclusively concentrate on the former statistic. The value of the X^2-statistic is evaluated against percentage points of the chi-square distribution with (I-1)(J-1) degrees of freedom. Given non-independence, one can inspect the residuals for specific patterns. While these patterns are easily visible in small tables, visually analysing more or less subtle relationships from a large table can become too difficult. In addition, the raw residuals themselves suffer from differences in size due to the differences in size of the original frequencies, and for comparing the residuals it is more appropriate to standardize them in some way. One obvious way is to use standardized residuals, which are equal to X_{ij}, the square root of the contribution of each cell to the X^2-statistic. A more subtle kind of standardization leads to Haberman's adjusted residuals (see e.g. Haberman, 1979). For three-way tables the situation is more complex as there are now three main effects, three two-way interactions and one three-way interaction

$$\log e_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$$

Again an unsaturated model consists of a subset of terms from the right hand side. A simple model is the three-way independence model consisting of u, $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$,, i.e .

$$\log e_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)},$$

or

$$e_{ijk} = f_{i++}f_{+j+}f_{++k}/n^2.$$

In this case we obtain the standardized residuals

$$X_{ijk} = (f_{ijk} - e_{ijk})/e_{ijk}^{1/2}.$$

As an example of a more complex model we may consider a model with the inclusion of one two-way interaction, say $\mathbf{u}_{13}$. It has the form

$$\log e_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)}$$

or,

$$e_{ijk} = f_{i+k}f_{+j+}/n.$$

Using this definition of e_{ijk}, we may define the standardized residuals as before.

A final model to be considered here, is the model which includes all two-way interactions but not the three-way interactions. For this model it is, however, not possible to formulate the expected values in closed form, and they have to be found via iteration, for example via the iterative proportional fitting algorithm (cf. Bishop et al., 1975, p. 83ff.). Again, the same formula applies for the standardized residuals.

3. SINGULAR VALUE DECOMPOSITION

3.1 Two-mode case

The singular value decomposition of arbitrary matrices has in the last decades become one of the work horses of data analysis and statistics (see e.g. Good, 1969, for a survey of applications using the singular value decomposition). If we take $\mathbf{Z}$ to be an arbitrary I×J matrix , the singular value decomposition is defined as

$\mathbf{Z} = \mathbf{AGB}'$ with $\mathbf{A}'\mathbf{A} = \mathbf{I}_I$, $\mathbf{B}'\mathbf{B} = \mathbf{I}_J$, and $\mathbf{G}$ diagonal.

In summation notation this may be written as

$$z_{ij} = \sum_p g_{pp}a_{ip}b_{jp}$$

The decomposition is not unique in the sense that any permutation of columns of $\mathbf{A}$ is allowed, provided the inverse permutation is applied to $\mathbf{G}$ and $\mathbf{B}$, and furthermore, if any of the g_{pp} are identical, there exists an invariant subspace for $\mathbf{A}$ and for $\mathbf{B}$. For the presentation here we will assume that all g_{pp} are different, and that they are arranged in descending order. With these side conditions the singular value decomposition of $\mathbf{Z}$ is uniquely defined. It can easily be shown that $\mathbf{A}$ and $\mathbf{B}$ are the canonical solutions or the eigenvector matrices of $\mathbf{ZZ}'$ and $\mathbf{Z}'\mathbf{Z}$ respectively, and that the singular values g_{pp} are the square roots of the eigenvalues. At most min (I,J) of the g_{pp} can be nonnegative. An important property of the singular value decomposition is that the best least squares approximation of $\mathbf{Z}_s$ of $\mathbf{Z}$ of rank s <

min(I,J), is equal to $\mathbf{Z}_S = \mathbf{A}_S\mathbf{G}_S\mathbf{B}_S'$ with $\mathbf{A}_S$ and $\mathbf{B}_S$ containing the first columns of **A** and **B** respectively, and with $\mathbf{G}_S$ the diagonal matrix with the s largest singular values.

3.2 Three-mode generalizations

If we assume that **Z** is an I×J×K three-mode array, there are several decompositions of **Z** which could claim to be the three-mode generalization of the singular value decomposition. We will discuss two of the more obvious ones.

Suppose we want to retain as many properties of the SVD as possible then the most simple decomposition is probably

$$\mathbf{Z} = \mathbf{AG}(\mathbf{C}'\otimes\mathbf{B}'),$$

or in summation notation

$$z_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b_{jq}c_{kr}g_{pqr}$$

with orthonormal **A**, **B**, and **C**, and **G** is restricted to be a super diagonal three-mode matrix, i.e. $g_{pqr} = 0$, unless p=q=r. A compacter way to write the above equation is

$$z_{ijk} = \sum_{s=1}^{S} a_{is}b_{js}c_{ks}g_{sss},$$

where S=P=Q=R, and the g_{sss} are the three-mode equivalents of the singular values, or generalized singular values. When discussing this model, we will always assume that S=P=Q=R.

In the psychometric literature this decomposition is known as the PARAFAC (Harshman, 1970; Harshman & Lundy, 1984) or the CANDECOMP model (Carroll & Chang, 1970), be it that generally no orthonormality constraints are imposed on **A**, **B**, and **C**. The matrices **A**, **B**, and **C** are not the canonical solutions i.e. the eigenvector matrices of $\Sigma_k\,\mathbf{Z}_k\mathbf{Z}_k'$, $\Sigma_i\,\mathbf{Z}_i\mathbf{Z}_i'$, and $\Sigma_j\,\mathbf{Z}_j\mathbf{Z}_j'$, respectively where $\mathbf{Z}_k = (z_{ij}^{k})$, $\mathbf{Z}_i = (z_{jk}^{i})$, and $\mathbf{Z}_j = (z_{ki}^{j})$. And in this sense the model is thus not a generalization of the SVD. In order to find an exact solution to the model (if it exists), or a least squares lower-rank approximation, one has to use a series of iterative conditional least squares problems (see Harshman & Lundy, 1984, for details). The restrictions placed on the matrix G make the orthonormal PARAFAC model rather difficult to fit, and the existence of an (approximate) solution is not guaranteed.

A second generalization of the singular value decomposition is generally known as the Tucker3 model (Tucker, 1966; Kroonenberg & De Leeuw, 1980; Kroonenberg, 1983). If an **exact** solution exists for this model, the **A**, **B**, and **C** are the canonical eigenvector matrices, but a simple structure of the singular values, g_{pqr}, is not postulated nor generally present. However, as in the orthonormal PARAFAC model, but in contrast with the two-mode case, the eigenvector matrices for the modes A, B, and C in a **lower-rank approximation** are no longer equal to the columns of the canonical solutions, and again an iterative conditional least squares algorithm has to be employed to find the estimates. However, in contrast with the orthonormal PARAFAC model, there always exists an approximate solution. The Tucker3 model can be written as

$$z_{ijk} = \sum_p \sum_q \sum_r a_{ip} b_{jq} c_{kr} g_{pqr}$$

with again orthonormal **A**, **B**, and **C**, but without restrictions on **G**. **G** is a three-way array of the order P×Q×R, containing the generalized singular values. In two-mode analysis and in the orthonormal PARAFAC model the singular values can always be chosen to be positive, however this is not the case in the present model, and elements of g_{pqr} may be negative. On the other hand, the squared generalized singular values in both mentioned three-mode models represent amounts of explained variation. The relationships between the components of modes A, B, and C are no longer one-to-one as in the previous model, and in principle all P×Q×R combinations of components may occur. Both three-mode models are unique in the same sense that the ordinary SVD is unique, given the orthonormality restrictions. In summary, not all properties of the two-mode singular value decomposition carry over to all three-mode models,and no three-mode model has all the properties of the two-mode SVD. A more detailed discussion of the issues examined in this section may be found in Chapter 1 of this volume, in particular in the papers by D'Aubigny, Denis-Dhorne, Franc, and Kruskal.

4. CORRESPONDENCE ANALYSIS FOR TWO-WAY TABLES

One of the aims of regular (two-mode) correspondence analysis is to portray the profile similarities between rows and/or columns of an I×J contingency table in an Euclidean space, generally a plane, in such a way that the rows (columns) which have similar conditional distributions given the marginal totals, are located close to each other.

Before defining the measure of similarity, it is necessary to introduce some notation. Let **F** be an I×J contingency table in which the elements f_{ij} indicate the number of times row category i and column category j occur together in a sample.

The vector of row totals is indicated by $\mathbf{f}_I = \{f_{i+}\}_{i=1,..,I}$, and the vector of column totals by $\mathbf{f}_J = \{f_{+j}\}_{j=1,..,J}$, and the total number of observations is $n = f_{++}$. In order to simplify the notation somewhat, we will couch the discussion in terms of relative frequencies with respect to n:

$$\mathbf{P} = \mathbf{F}/n;\ p_{ij} = f_{ij}/n;\ \mathbf{p}_I = \{p_{i+}\};\ \text{and}\ \mathbf{p}_J = \{p_{+j}\}.$$

Furthermore, we will use $\mathbf{D}_I$ and $\mathbf{D}_J$ to indicate the diagonal matrices which have $\mathbf{p}_I$ and $\mathbf{p}_J$ on their diagonals, respectively. A **profile** of a row i is the vector of values p_{ij}/p_{i+}, which can be interpreted as conditional proportions.

The measure of similarity, called χ^2-distance , between rows i and i' of the table **P** is defined as (see e.g. Benzécri, 1976; Gifi, 1981, Ch. 3)

$$d^2_{ii'} = \sum_{j=1}^{J} \left(\frac{p_{ij}}{p_{i+}} - \frac{p_{i'j}}{p_{i'+}} \right)^2 / p_{+j} = \sum_{j=1}^{J} (h_{ij} - h_{i'j})^2,$$

where

$$h_{ij} = (p_{ij}/p_{i+})/p_{+j}^{1/2}.$$

Thus the χ^2-distance between rows i and i' is a measure for the difference between the profiles of rows i and i' (see also Van der Heijden, 1987, p.29ff). From the definition , we see that the χ^2-distance between row i and i' of **P** is the same as the Euclidean distance between rows i and i' of $\mathbf{H} = \{h_{ij}\}$, which can be expressed in matrix notation as

$$\mathbf{H} = \mathbf{D}_I^{-1}\mathbf{P}\mathbf{D}_J^{-1/2}.$$

In order to find a representation of the rows in Euclidean space, we have to search for some $\mathbf{Y}_r$, such that $\mathbf{Y}_r\mathbf{Y}_r' = \mathbf{H}\mathbf{H}'$. It is desirable that such an $\mathbf{Y}_r$ is chosen which optimizes, for instance, the explained variation in **H**, so that in practical examples one may settle for a low dimensional subspace of R^I. The standard approach to this problem with the desired optimizing properties is via the singular value decomposition (SVD) of a matrix. Instead of directly defining the SVD of **H**, it is advantageous, as will be shown, to find the singular vectors as follows

$$\mathbf{H} = \mathbf{D}_I^{-1}\mathbf{P}\mathbf{D}_J^{-1/2} = \mathbf{D}_I^{-1/2}(\mathbf{D}_I^{-1/2}\mathbf{P}\mathbf{D}_J^{-1/2}) = \mathbf{D}_I^{-1/2}(\hat{\mathbf{A}}\hat{\mathbf{G}}\hat{\mathbf{B}}')$$

with $\hat{\mathbf{A}}'\hat{\mathbf{A}} = \mathbf{I}_I$, $\hat{\mathbf{B}}'\hat{\mathbf{B}} = \mathbf{I}_J$, and $\hat{\mathbf{G}}$ the diagonal matrix with singular values. The desired representation of rows in R^I may be taken as

$$\mathbf{Y}_r = \mathbf{D}_I^{-1/2}\hat{\mathbf{A}}\hat{\mathbf{G}}$$

The appropriate normalization for $\mathbf{Y}_r$ is $\mathbf{Y}_r'\mathbf{D}_I\mathbf{Y}_r = \hat{\mathbf{G}}\hat{\mathbf{A}}'\mathbf{D}_I^{-1/2}\mathbf{D}_I\mathbf{D}_I^{-1/2}\hat{\mathbf{A}}\hat{\mathbf{G}} = \hat{\mathbf{G}}^2$.

This metric is extensively discussed in for instance Caillez and Pagès (1976). It can be shown that indeed $\mathbf{Y}_r\mathbf{Y}_r' = \mathbf{HH}'$, that the first singular value is 1, and that the first column of $\mathbf{Y}_r$ consists of ones (see e.g. Gifi, 1981, p. 136). As the first column does not contribute to the distances between rows, it may be eliminated by redefining $\mathbf{H}$ as

$$\mathbf{H}^* = \mathbf{H} - \mathbf{uu}'\mathbf{D}_J^{1/2},$$

where $\mathbf{u}$ is a column vector of ones. In this way the first singular vectors are eliminated so that

$$\mathbf{H}^* = \mathbf{D}_I^{-1/2}(\mathbf{D}_I^{-1/2}\mathbf{P}\mathbf{D}_J^{-1/2} - \mathbf{D}_I^{1/2}\mathbf{uu}'\mathbf{D}_J^{1/2}) = \mathbf{D}_I^{-1/2}\mathbf{AGB}'$$

where $\mathbf{A}$ ($\mathbf{B}$) contains the non-trivial singular vectors of $\hat{\mathbf{A}}$($\hat{\mathbf{B}}$). Similarly $\mathbf{Y}_r$ is redefined as

$$\mathbf{Y}_r = \mathbf{D}_I^{-1/2}\mathbf{AG}$$

If we define $\mathbf{X}$ to be the expression in brackets above, we see that

$$x_{ij} = p_{ij}/(p_{i+}p_{+j})^{1/2} - (p_{i+}p_{+j})^{1/2} = (p_{ij} - p_{i+}p_{+j})/(p_{i+}p_{+j})^{1/2} =$$
$$= n^{-1/2}(f_{ij} - e_{ij})/e_{ij}^{1/2},$$

where the e_{ij} is the expected value for the cell (i,j) of $\mathbf{F}$ under the null model of independence of row and column classifications. Thus we see that the SVD is computed for the matrix of standardized residuals disregarding the factor $n^{-1/2}$, .

In order to investigate the association between row and column classifications, it is useful to find a representation $\mathbf{Y}_c^*$ for the columns of the table with a similar normalization to $\mathbf{Y}_r$, such a representation is

$$\mathbf{Y}_c^* = \mathbf{D}_J^{-1/2}\hat{\mathbf{B}}, \quad \text{or} \quad \mathbf{Y}_c^* = \mathbf{D}_J^{-1/2}\mathbf{B}$$

depending on whether the trivial singular vector has been eliminated or not . The normalization gives $\mathbf{Y}_c^{*\prime}\mathbf{D}_J\mathbf{Y}_c^* = \mathbf{B}'\mathbf{D}_J^{-1/2}\mathbf{D}_J\mathbf{D}_J^{-1/2}\mathbf{B} = \mathbf{I}$. The relationship between $\mathbf{Y}_r$ and $\mathbf{Y}_c^*$ then follows from

$$\mathbf{D_I^{-1}PY_c^*} = \mathbf{D_I^{-1/2}(D_I^{-1/2}PD_J^{-1/2})B} = \mathbf{D_I^{-1/2}(AGB' + D_I^{1/2}uu'D_J^{1/2})B}$$

$$= \mathbf{D_I^{-1/2}AG + uu'D_J^{-1/2}B} = \mathbf{D_I^{-1/2}AG} = \mathbf{Y_r},$$

as the columns of **B** are orthogonal to **u**. The importance of the relationship $\mathbf{D_I^{-1}PY_c^*} = \mathbf{Y_r}$, and thus of the chosen representations of $\mathbf{Y_r}$ and $\mathbf{Y_c^*}$ is that a row category (point) is the centre of gravity of the column points, when the latter are weighted by the frequency of the rows . This is called the "barycentric principle" (Benzécri, 1976).

The above results were derived for the rows , and the parallel results can be derived for the columns by interchanging the roles of rows and columns. Such an interchange leads to $\mathbf{Y_c} = \mathbf{D_J^{-1/2}BG}$, and $\mathbf{Y_r^*} = \mathbf{D_I^{-1/2}A}$, and again the barycentric principle holds.

The procedure to find the representations for the rows and columns for which the Euclidean distances between rows (columns) correspond to dissimilarities (expressed as χ^2-distances) between the rows (columns) of a contingency table may be summarized as follows.

1. Determine **X**, where **X** is the matrix of standardized residuals from the model postulating independence between row and column classifications
2. Determine the SVD, **AGB'**, of **X**
3. Define $\mathbf{Y_r} = \mathbf{D_I^{-1/2}AG}$ and $\mathbf{Y_c^*} = \mathbf{D_J^{-1/2}B}$, so that $\mathbf{D_I^{-1}P\,Y_c^*} = \mathbf{Y_r}$, and the barycentric principle holds for the row points.

   Define $\mathbf{Y_c} = \mathbf{D_J^{-1/2}BG}$ and $\mathbf{Y_r^*} = \mathbf{D_I^{-1/2}A}$, so that $\mathbf{D_J^{-1}P'\,Y_r^*} = \mathbf{Y_c}$, and the barycentric principle holds for the column points.
4. Plot $\mathbf{Y_r}$ and $\mathbf{Y_c}$ in the same graph

Plotting as suggested in Step 4 is the common procedure in the French school of correspondence analysis, and it gives a symmetric display of row and column points in the same scale, furthermore supplying an interpretation of the distances between the rows and columns, respectively, but it does not provide as direct an interpretation of column and row distances (see also Novak and Hoffman, 1987, p.12, 13, 16). Or as Deville and Saporta (1983) point out the simultaneous

representation is merely a device, and has no strong theoretical background because the categories of the rows and columns belong to two different vector spaces.

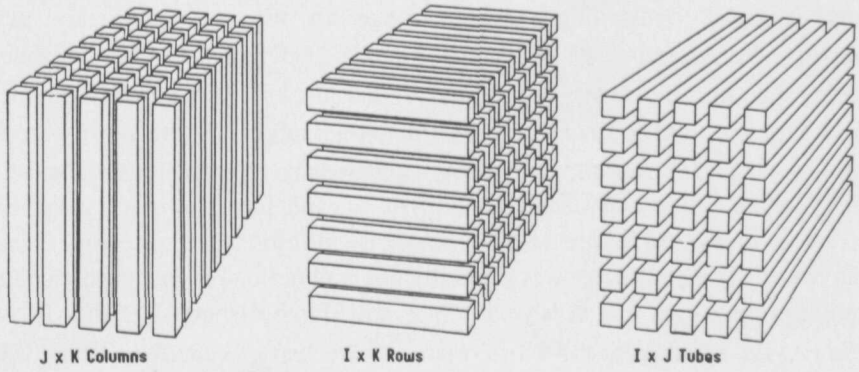

Figure 1: *Fibers: One-way submatrices of a three-way matrix*

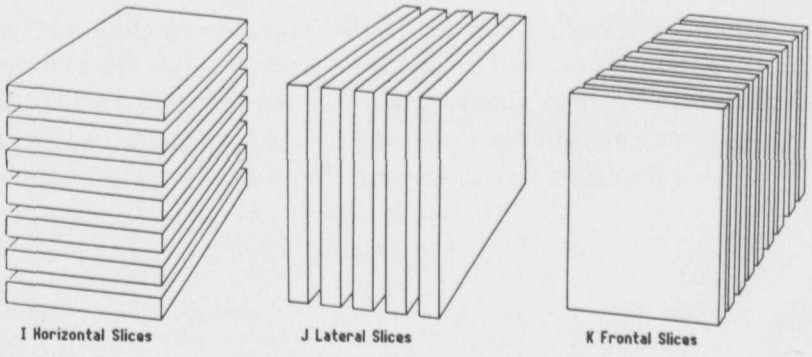

Figure 2. *Slices: Two-way submatrices of a three-way matrix*

5. CORRESPONDENCE ANALYSIS FOR THREE-WAY TABLES

The core of the extension of correspondence analysis to three-way tables is the generalization of the procedure described in the previous section using various forms of three-mode procedures as generalizations of the singular value decomposition, and using generalizations of the χ^2-distance to define the distances between profiles in a three-way table.

Definitions. Let **F** be an I×J×K contingency table; the element f_{ijk} indicates the

number of times category i, category j and category k occur together. The totals for each i are f_{i++}, the totals for each j f_{+j+}, and totals for each k f_{++k}; $n = f_{+++}$ is the total number of observations. Again we define porportions relative to n: $\mathbf{P} = \mathbf{F}/n$, $p_{ijk} = f_{ijk}/n$, $p_{i++} = f_{i++}/n$, $p_{+j+} = f_{+j+}/n$, and $p_{++k} = f_{++k}/n$. $\mathbf{D}_I$, $\mathbf{D}_J$, and $\mathbf{D}_K$ are diagonal matrices containing on their diagonals the p_{i++}, p_{+j+}, and p_{++k}, respectively. Furthermore, we define $\mathbf{D}_{IJ}$ as the IJ×IJ diagonal matrix with the p_{ij+} on its diagonal, or a block diagonal matrix with J blocks of I×I diagonal matrices; $\mathbf{D}_{IK}$ and $\mathbf{D}_{JK}$ are similarly defined. A three-way table may be represented as collections of rows, columns, and tubes - generically referred to as **fibers** (see Figure 1), or as collections of horizontal, lateral, and frontal slices - generically referred to as **slices** (see Figure 2). The (implicit) definitions are necessary because the concept of, for instance, rows is generally not well defined in the present context. In formulae a three-way matrix is generally assumed to be arranged as an I×JK two-way matrix, i.e. effectively the I J×K-matrices are juxtaposed: $\mathbf{P} = (\mathbf{P}_1, \mathbf{P}_2, .., \mathbf{P}_I)$. Other permutations of the indices occur as well; which one will be clear from the context.

χ^2-distances. For two-way contingency tables the χ^2-distance is used to define the dissimilarities between profiles, and the definition of distance is the same for columns and rows. In this paper we approach the χ^2-distance in the same way for a three-way table, i.e. one single definition should be used for the three ways of the table. Treating the three ways even handed, there seem to be two different ways in which one could define such χ^2-distances, based on two different ways to construct profiles. In the first option, profiles are defined for slices, and in the second option profiles are defined for fibers, respectively referred to as **slice profiles** and **fiber profiles**

5.1 Slice profiles

The χ^2-distance between horizontal slice i and slice i' may be defined as

$$\partial^2_{ii'} = \sum_{j=1}^{J} \sum_{k=1}^{K} \{ h_{ijk} - h_{i'jk} \}^2,$$

with

$$h_{ijk} = (p_{ijk}/p_{i++})/p_{+j+}^{1/2} p_{++k}^{1/2},$$

and thus **H** has the following matrix form

$$\mathbf{H} = \mathbf{D}_I^{-1}\mathbf{P}(\mathbf{D}_K^{-1/2} \otimes \mathbf{D}_J^{-1/2}).$$

The implication in this case is that we seek to find a representation of the horizontal slices in a Euclidean space such that slices which have similar profiles are located close to one another. Thus we want to find a representation $\mathbf{Y}_h$ of $\mathbf{H}$ such that $\mathbf{Y}_h\mathbf{Y}_h' = \mathbf{H}\mathbf{H}'$. Analogously to the two-way case described above we may proceed to rewrite $\mathbf{H}$

$$\mathbf{H} = \mathbf{D}_I^{-1/2}\left[\mathbf{D}_I^{-1/2}\,\mathbf{P}(\mathbf{D}_K^{-1/2} \otimes \mathbf{D}_J^{-1/2})\right] = \mathbf{D}_I^{-1/2}\hat{\mathbf{A}}\hat{\mathbf{G}}(\hat{\mathbf{C}}' \otimes \hat{\mathbf{B}}')$$

with orthonormal $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, and $\hat{\mathbf{C}}$. For the orthonormal PARAFAC model , $\hat{\mathbf{G}}$ is equal to $\hat{\Delta}\mathbb{I}$, with $\mathbb{I}$ the super diagonal three-way identity matrix (i.e. $i_{pqr} = 0$, unless p=q=r; $i_{pqr} = 1$, if p=q=r) , and $\hat{\Delta}$ the diagonal matrix with generalized singular values. For the Tucker3 model $\hat{\mathbf{G}}$ is unrestricted. The Euclidean representations for the horizontal, lateral, and frontal slices become

$$\mathbf{Y}_h = \mathbf{D}_I^{-1/2}\hat{\mathbf{A}}\hat{\mathbf{G}} \quad (= \mathbf{D}_I^{-1/2}\hat{\mathbf{A}}\hat{\Delta}\mathbb{I}; \text{ for the PARAFAC model})$$

$$\mathbf{Y}_l^* = \mathbf{D}_J^{-1/2}\hat{\mathbf{B}}$$

$$\mathbf{Y}_f^* = \mathbf{D}_K^{-1/2}\hat{\mathbf{C}}.$$

Note, that $\mathbf{Y}_h$ is an I×(J×K) matrix, or in case of an approximate lower-rank solution an I×(Q×R) matrix. $\mathbf{Y}_l^*$ is an J×J or J×Q matrix, and $\mathbf{Y}_f^*$ a K×K or K×R matrix. Thus the representation of the horizontal slices is of the order Q×R in the Tucker3 case. In the orthonormal PARAFAC model $\mathbf{G} = \hat{\Delta}\mathbb{I}$, and only those columns for which q=r have non-zero elements, so that there are s=p=q=r non-zero columns. In other words, the representations of all three modes have the same number of columns for this model.

From the above equations it can be verified that indeed $\mathbf{Y}_h\mathbf{Y}_h' = \mathbf{H}\mathbf{H}'$, and similar arguments as in ordinary correspondence analysis (see De Leeuw, 1983, p. 129, which paper also includes other references) show that the first singular value is 1 and that $\mathbf{Y}_h$ has a unit first column. Thus again we can eliminate trivial singular vectors by redefining $\mathbf{H}$

$$\mathbf{H}^* = \mathbf{H} - \mathbf{u}(\mathbf{u}' \otimes \mathbf{u}')(\mathbf{D}_K^{1/2} \otimes \mathbf{D}_J^{1/2})$$

and

$$\mathbf{H}^* = \mathbf{D}_I^{-1/2}\left[\mathbf{D}_I^{-1/2}\mathbf{P}(\mathbf{D}_K^{-1/2} \otimes \mathbf{D}_J^{-1/2}) - \mathbf{D}_I^{1/2}\mathbf{u}(\mathbf{u}' \otimes \mathbf{u}')(\mathbf{D}_K^{1/2} \otimes \mathbf{D}_J^{1/2})\right]$$

$$= \mathbf{D}_I^{-1/2}\mathbf{X} = \mathbf{D}_I^{-1/2}\mathbf{AG}(\mathbf{C}' \otimes \mathbf{B}'),$$

in which $\mathbf{X}$ is decomposed with a three-mode model or a generalized singular value decomposition without the first singular vectors or components. The representations are accordingly adjusted to $\mathbf{Y}_h = \mathbf{D}_I^{-1/2}\mathbf{AG}$, $\mathbf{Y}_l^* = \mathbf{D}_J^{-1/2}\mathbf{B}$, and $\mathbf{Y}_f^* = \mathbf{D}_K^{-1/2}\mathbf{C}$.

The elements of $\mathbf{X}$ can be written as

$$\begin{aligned} x_{ijk} &= p_{ijk}/(p_{i++}p_{+j+}p_{++k})^{1/2} - (p_{i++}p_{+j+}p_{++k})^{1/2} \\ &= (p_{ijk} - p_{i++}p_{+j+}p_{++k})/(p_{i++}p_{+j+}p_{++k})^{1/2} \\ &= n^{-1/2}(f_{ijk} - e_{ijk})/e_{ijk}^{1/2}, \end{aligned}$$

where $e_{ijk} = f_{i++}f_{+j+}f_{++k}/n^2$ is equal to the expected value of f_{ijk} under the null hypothesis of independence of the three classifications. With $\mathbf{Y}_h$, $\mathbf{Y}_l^*$, and $\mathbf{Y}_f^*$ defined as above, the barycentric principle holds in a special way

$$\mathbf{D}_I^{-1}\mathbf{P}(\mathbf{Y}_f^* \otimes \mathbf{Y}_l^*) = \mathbf{D}_I^{-1}\mathbf{P}(\mathbf{D}_K^{-1/2} \otimes \mathbf{D}_J^{-1/2})(\mathbf{C} \otimes \mathbf{B}) =$$

$$= \mathbf{D}_I^{-1/2}\left[\mathbf{D}_I^{-1/2}\mathbf{P}(\mathbf{D}_K^{-1/2} \otimes \mathbf{D}_J^{-1/2})(\mathbf{C} \otimes \mathbf{B})\right] =$$

$$= \mathbf{D}_I^{-1/2}\left[\mathbf{AG}(\mathbf{C}' \otimes \mathbf{B}') + \mathbf{D}_I^{1/2}\mathbf{u}(\mathbf{u}' \otimes \mathbf{u}')(\mathbf{D}_K^{1/2} \otimes \mathbf{D}_J^{1/2})\right](\mathbf{C} \otimes \mathbf{B}) =$$

$$= \mathbf{D}_I^{-1}\mathbf{AG} + \mathbf{u}(\mathbf{u}' \otimes \mathbf{u}')(\mathbf{D}_K^{1/2} \otimes \mathbf{D}_J^{1/2})(\mathbf{C} \otimes \mathbf{B}) =$$

$$= \mathbf{D}_I^{-1}\mathbf{AG} = \mathbf{Y}_h,$$

with

$$y^h_{i(qr)} = p_{i++}^{-1/2} \sum_{p=1}^{P} a_{ip}g_{pqr} = \sum_{j=1}^{J}\sum_{k=1}^{K}(p_{ijk}/p_{i++})b_{jq}c_{kr}.$$

Thus the coordinates of the horizontal slices are at the centre of gravity of all lateral and frontal slice combinations as expressed in the reduced spaces. If the orthonormal PARAFAC model has been fitted to X then the barycentric principle is only relevant for those columns $y^h_{i(qr)}$ for which q=r, as the other columns consist

only of zero elements.

Without going into the details of the derivations, the symmetry of the problem and its solution (slightly obscured by the Kronecker notation) assure that when the same χ^2-distance is used for the frontal slices and for the lateral slices, the same matrix of standardized residuals from the three-way independence model results. And thus also the same type of representations $\mathbf{Y}_I$ and $\mathbf{Y}_f$ are derived. In other words, defining the χ^2-distance for slices leads to a three-way generalization of correspondence analysis which is consistent and symmetrical in its treatment of the three modes. The solution is found in a similar way as in the two-mode case, i.e. by fitting the three-way independence model, calculating the standardized residuals, decomposing the residuals via a three-way generalization of the SVD, and scaling the singular vectors in the way derived above. To stay in line with the two-mode procedures $\mathbf{Y}_h$, $\mathbf{Y}_I$ and $\mathbf{Y}_f$ should be used for plotting, i.e. $\mathbf{Y}_h = \mathbf{D}_I^{-1/2}\mathbf{AG}$, $\mathbf{Y}_I = \mathbf{D}_J^{-1/2}\mathbf{BG}$, and $\mathbf{Y}_f = \mathbf{D}_K^{-1/2}\mathbf{CG}$ should be used , where $\mathbf{G}$ is written as a P×(Q×R), Q×(R×P), and a R×(P×Q) matrix, respectively.

In the case of the orthonormal PARAFAC model, the common practice described above for two-way correspondence analysis for plotting representations can be followed, i.e. $\mathbf{Y}_h$,$\mathbf{Y}_I$, and $\mathbf{Y}_f$ can directly be displayed in one single plot, because of the common dimensionality and generalized singular values (i.e. the g_{sss}). As we remarked above in this way the χ^2-distances for each of the three types of slices are portrayed in the plot through the Euclidean distances between the points within a mode; not so easily interpreted are the between-mode distances, parallel to the situation of ordinary correspondence analysis. How the representations may be plotted simultaneously in the case of the Tucker3 model is all but clear, as each of the representations has a different order (unless P=Q=R), and the only thing the representations share, is a common sum of squared singular values. In this paper we will skirt the issue of plotting representations from the Tucker3 model.

5.2 Fiber profiles

The development of the formulas for the fiber profiles follows largely the pattern of the previous section, and therefore, the derivations will not be presented in great detail.

The χ^2-distance between fibers, here rows ik and i'k', can be defined as

$$\delta^2_{(ik)(i'k')} = \sum_{j=1}^{J} (h_{ijk} - h_{i'jk'})^2$$

with

$$h_{ijk} = (p_{ijk}/p_{i+k})/p_{+j+}^{1/2},$$

or

$$\mathbf{H} = \mathbf{D}_{IK}^{-1}\mathbf{P}\mathbf{D}_J^{-1/2}.$$

Analogously to the previous section, this may be decomposed as

$$\mathbf{H} = \mathbf{D}_{IK}^{-1/2}(\mathbf{D}_{IK}^{-1/2}\mathbf{P}\mathbf{D}_J^{-1/2}) = \mathbf{D}_{IK}^{-1/2}(\hat{\mathbf{C}} \otimes \hat{\mathbf{A}})\hat{\mathbf{G}}\hat{\mathbf{B}}',$$

and

$$\mathbf{H}^* = \mathbf{D}_{IK}^{-1/2}\left[\mathbf{D}_{IK}^{-1/2}\mathbf{P}\mathbf{D}_J^{-1/2} - \mathbf{D}_{IK}^{1/2}(\mathbf{u}\otimes\mathbf{u})\mathbf{u}'\mathbf{D}_J^{1/2}\right] = \mathbf{D}_{IK}^{-1/2}\mathbf{X} =$$

$$= \mathbf{D}_{IK}^{-1/2}(\mathbf{C} \otimes \mathbf{A})\mathbf{G}\mathbf{B}'.$$

This leads to representations $\mathbf{Y}_r = \mathbf{D}_{IK}^{-1/2}(\mathbf{C} \otimes \mathbf{A})\mathbf{G}$ and $\mathbf{Y}_l^* = \mathbf{D}_J^{-1/2}\mathbf{B}$, and they can be shown to satisfy the barycentric principle. The difference between these representations and those in the previous section is that we have a mixed fiber and slice representation. Before, due to the decomposition of $\mathbf{D}_{IK}$ into $\mathbf{D}_K \otimes \mathbf{D}_I$, we managed to separate the first and third mode representations. A clearer view of this may be had by writing out the matrix $\mathbf{X}$

$$X_{ijk} = p_{ijk}/(p_{i+k}p_{+j+})^{1/2} - (p_{i+k}p_{+j+})^{1/2} =$$

$$= (p_{ijk} - p_{i+k}p_{+j+})/(p_{i+k}p_{+j+})^{1/2} = n^{-1/2}(f_{ijk} - e_{ijk})/e_{ijk}^{1/2},$$

so that $\mathbf{X}$ turns out to be the matrix with standardized residuals from the loglinear model including the $\mathbf{u}_{13}$-term. Similarly, for **tubes** we end up with $\mathbf{X}$ as the matrix of standardized residuals of the loglinear model including the $\mathbf{u}_{12}$-term, and for columns the loglinear model includes the term $\mathbf{u}_{23}$. In other words, the χ^2-distances for rows, columns, and tubes give rise to different standardized residuals, and thus different sets of generalized singular vectors. This non-symmetric treatment can be seen as a disadvantage in the sense that no one single coherent solution emerges; it can also be seen as an advantage, in the sense that the residuals of three different loglinear models can be analysed.

Given this situation, one might wonder whether there would exist a definition of the χ^2-distance that would allow the analysis of standardized residuals of loglinear

models with two two-way interaction terms. However, this does not seem to be possible without additional assumptions or different approaches. The reason for this is that first of all there seem to exist only two definitions of profiles in a three-way table, and, in addition, the matrix **X** has the form

$$X_{ijk} = p_{ijk}/(p_{i+k}p_{+jk})^{1/2} - (p_{i+k}p_{+jk})^{1/2}.$$

In this expression i, j, and k are involved in two two-dimensional margins, and it seems difficult to define what constitutes a profile in this context. The situation becomes even more difficult if one would like to consider profiles and χ^2-distances which lead to the decomposition of standardized residuals from the loglinear model with all three two-way interactions included, because in that case, no explicit formulae exists for the expected values of this model. This seems to make developments as the above entirely impossible.

6. DISCUSSION AND CONCLUSION

The conceptually most satisfying generalization of correspondence analysis to three-way data seems to be via χ^2-distances defined on slices, and the decomposition of the standardized residuals (of the three-way independence model) with the orthonormal PARAFAC model. It leads to a possibility of simultaneously plotting all three modes, a symmetric representation of the three classifications, and relatively simple formulae, which bear a great resemblance to the two-mode case. The greatest problem is that the orthonormal PARAFAC does not always have an approximate solution. How serious this is in practice remains to be seen. In particular, if only one component for each mode A, B, and C is necessary (i.e. S=P=Q=R=1) no problem exists, because such an approximate solution does always exist. In this case the two models mentioned, the Tucker3 model and the (orthonormal) PARAFAC model, are equivalent.

Theoretically, the generalizations do not seem very succesful if they are viewed from the vantage point of unravelling interactions from loglinear models. The methods developed by Van der Heijden and De Leeuw (1985; Van der Heijden, 1987, this volume) seem to be more adequate for that purpose. A further reasonable alternative seems to be the non-correspondence-analysis route taken by Kroonenberg (1983) to directly decompose the standardized residuals from arbitrary loglinear model by three-mode methods. In particular, this allows an orthogonal partitioning of X^2 (the sum of the squared standardized residuals) by multiplicative terms. Within this light, the above method (χ^2-metric on slices) is just a special case, with the additional justification of being derived from "first principles", i.e. Euclidean representation of profiles and the relationship between correspondence analysis and the singular value decomposition. Further explorations

into this approach, especially with different three-mode methods are in progress.

REFERENCES

Benzécri, J.-P. (1976). *L'Analyse des données. II: Analyse des correspondances* [Data analysis. II: Correspondence analysis]. Paris: Dunod.

Bishop, Y.Y.M., Fienberg, S.E., and Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge, MA: The MIT press.

Carroll, J.D., and Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika,* **35**, 283-319.

Cai llez, F., and Pagès, J.P. (1976). *Introduction à l'analyse des données* [Introduction to data analysis]. Paris: SMASH.

De Leeuw, J. (1983). Models and methods for the analysis of correlation matrices.*Journal of Econometrics,* **22**, 113-137.

Deville,J.-C., and Saporta, G. (1983). Correspondence analysis with an extension towards nominal time series. *Journal of Econometrics,* **22**,.169-189.

Gifi, A. (1981). *Nonlinear multivariate analysis.* (Preliminary edition). Leiden: Department of Data Theory.

Good, I.J. (1969). Some applications of the singular value decomposition. *Technometrics,* **11**, 823-831.

Haberman, S.J. (1979). *Analysis of qualitative data.* New York: Academic Press.

Harshman, R.A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-mode factor analysis. *UCLA Working Papers in Phonetics,* **16**, 1-84 (Reprinted by Xerox University Microfilms, Ann Arbor, MI, order no. 10,085)

Harshman, R.A., and Lundy, M.E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H.G. Law, C.W. Snyder Jr., J.A. Hattie, R.P.McDonald (Eds.) *Research methods for multimode data analysis* (pp.122-215). New York: Praeger.

Kroonenberg, P.M. (1983). *Three-mode principal component analysis: Theory and applications.* Leiden: DSWO Press.

Kroonenberg, P.M., and De Leeuw, J. (1980). Principal component analysis for three-mode data by means of alternating least squares algorithms. *Psychometrika,* **45**, 69-97.

Novak, T.P., and Hoffman, D.L. (1987). *Residual scaling using the singular value decomposition: Graphical representation of log-linear models.* (Research working paper No. 87-1). Columbia Business School, New York.

Van der Heijden, P.G.M. (1987). *Correspondence analyis of longitudinal categorical data..* Leiden: DSWO Press.

Van der Heijden, P.G.M., and De Leeuw, J. (1985). Correspondence analysis used complementary to loglinear analysis. *Psychometrika,* **50** , 429-447.

Tucker, L.R (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika,* **31**, 279-311.

Acknowledgement: Thanks are due to Piet Brouwer for permission to use his drawings for Figures 1 and 2